

Dopant Fluctuations and Quantum Effects in Sub-0.1 μ m CMOS

Chris Bowen, Gerhard Klimeck, Mike Goodwin and Dick Chapman
Corporate R&D, Texas Instruments Incorporated, Dallas, TX 75235

I. INTRODUCTION

The scaling of CMOS gate lengths below 0.1 μ m will expose physical phenomena which are not important for current (and past) technologies. Among these are the discrete nature of dopant ions, tunneling, and size quantization. In this work we attempt to quantify the magnitude of these effects and identify methods to reduce or eliminate device degradation due to them.

To date, transistor dimensions are much larger than the screening length of an individual dopant ion ($\approx 100\text{\AA}$). Moreover, the number of dopant ions in a single device is rather large (> 1000). Given these conditions it is perfectly reasonable to assume that the net effect of the dopant ions may be described by a continuous charge distribution. Indeed, this assumption has proven to be quite accurate in device simulation for current and past technology nodes. As we scale the gate length below 0.1 μ m, however, we must begin questioning the validity this assumption. Another common assumption is that the current flowing through the gate insulator is negligible and can be ignored. Sub-0.1 μ m technologies will require¹ effective gate oxide thicknesses below 30 \AA . Since direct tunneling current increases exponentially with reduced oxide thickness, a limit will be reached where the thickness can no longer be scaled. Finally, degradation of transistor performance due to carrier confinement quantization should be investigated for very small geometry structures. We shall investigate these issues using device simulation tools and suggest methods to reduce or avoid their effects.

II. DOPANT ION FLUCTUATIONS

As previously mentioned the discrete nature of dopant ion charge will become evident in scaled technologies.^{2,3} We have performed a simulation study on the magnitude of the effect of channel dopant ion position fluctuations in scaled MOS devices. A three-dimensional drift-diffusion simulator (TMA Davinci) is employed to simulate carrier transport. Since we are using a drift-diffusion model we cannot (with any certainty) make any conclusions about scattering-limited transport regimes. We therefore limit our analysis to the MOSFET sub-threshold region.

The charge due to a dopant ion is modelled by assigning a meshpoint with a dopant concentration equal to the inverse of the volume associated with that meshpoint. In our simulations a uniform grid spacing of 2nm is used in the channel region. The test device used in the analysis is an NMOS with $t_{OX} = 3nm$, $x_j = 10nm$ and $N_{CHAN} = 1 \times 10^{18} cm^{-3}$. In Fig. 1 the sub-threshold potential barrier for electrons injected from the source is illustrated for a $W = L = 40nm$ device.

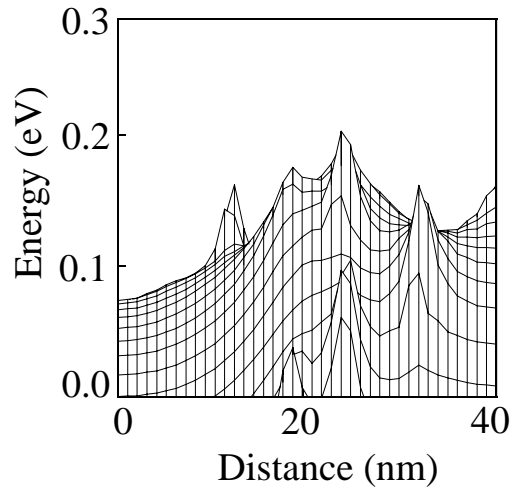


FIG. 1. Electron potential barrier for source electrons in a $W = L = 40nm$ NMOS for $V_{GS} = 0.3V$ and $V_{DS} = 1.0V$ (subthreshold region).

The location and magnitude of the “peaks” and “valleys” in the source barrier will depend on the arrangement of the channel dopant ions. In Fig. 2, $I_{DS}-V_{GS}$ characteristics for different random ion arrangements are shown. The dots in this figure correspond to the continuous dopant charge simulation. Note the discrete dopant ion simulations predict a significant variation in V_T as well as a negative shift in the average V_T compared with the continuous simulation. The negative shift is due to the fact that thermal transport over a barrier decreases exponentially with energy. Since the source barrier due to discrete ions fluctuates (Fig. 1), a disproportionate amount of the current flows through a potential “valley” thus reducing V_T .

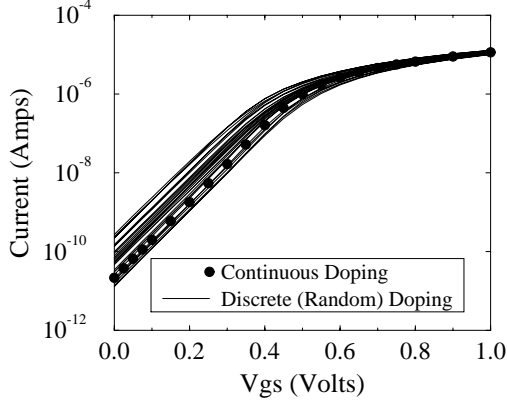


FIG. 2. I_{DS} vs. V_{GS} for continuous (dots) and discrete random (lines) channel doping distributions. The discrete random profiles exhibit significant variations in threshold voltage depending on the arrangement of channel dopant ions.

It is the variation in V_T due to ion arrangement that is of concern for scaled technologies. The standard deviation in sub-threshold V_T (voltage at $I_{DS} = 1nA$) for 100 random dopant distributions is $38mV$ for the $W = L = 40nm$ device. In order to understand the dependence of this variation on device geometry, simulations were carried out for three different gate lengths and two different gate widths. The results are summarized in Table I. As expected, the variation in threshold decreases significantly as the gate length increases. However, these simulations indicate that threshold variance does not decrease for wider devices. The reason for this is that the subthreshold current flow is highly localized. Nearly all of the current flows through the path with the lowest barrier height.

W	L	σV_T
40nm	40nm	38mV
40nm	80nm	27mV
40nm	120nm	20mV
80nm	40nm	39mV

TABLE I. Threshold voltage ($I_{DS} = 1nA$) standard deviation due to random dopant ion arrangement for various gate lengths and widths. Sample size for all cases is 100. Variation in threshold voltage decreases as gate length increases. However, there is no dependence on gate width. This is due to the fact that the sub-threshold current flow is highly localized.

From this simulation study it is evident that random dopant ion arrangement is a formidable issue for sub-0.1 μm CMOS. What can be done to reduce or eliminate these effects? The most obvious solution is to remove the dopant ions from the vicinity of the current flow path. To test this, channel dopant ions were completely removed from the surface ($0 - 5nm$) of the $L = W = 40nm$ structure. Indeed, this results in a reduction in threshold standard deviation of 32% (from $38mV$ to $26mV$). However, it will be hard to push such techniques much further. The reason for this is dopant ion charge will always be required to screen the drain potential from the source. An ultimate solution is provided by a structure composed of a narrow semiconductor channel sandwiched between gates on top and bottom (see insert of Fig. 3). In this structure, the drain potential can be screened by charge on the opposing gates provided that the silicon film is sufficiently thin (i.e. $\approx L_G/4$).^{4,5} Therefore, no dopant ions are required in the channel. In Fig. 3, simulated I_{DS} - V_{GS} curves for a double gate device are shown for varying channel doping. For channel doping levels below $1 \times 10^{17}cm^{-3}$ the I_{DS} - V_{GS} characteristic is independent of doping level and thus impervious to dopant fluctuations.

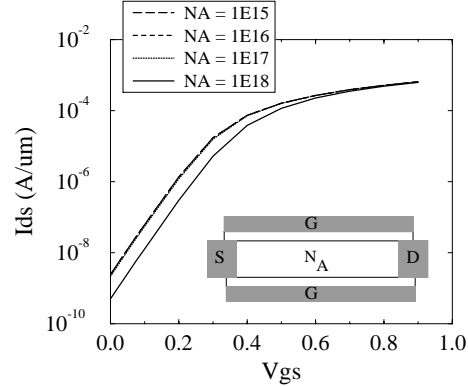


FIG. 3. I_{DS} - V_{GS} curves for a double gate device with $t_{OX} = 30\text{\AA}$, $L_G = 40nm$ and $t_{Si} = 8nm$ for different channel doping concentrations. The IV characteristics remain unchanged for doping concentrations below $1 \times 10^{17}cm^{-3}$. The drain potential is completely screened by the opposing gates so no dopant ions are required in the channel.

III. GATE INSULATOR TUNNELING

As the gate length of the MOS transistor is scaled, the gate oxide thickness must also be scaled.¹ The ultimate scaling limit for gate oxide thickness will most likely be determined by direct tunneling current. To estimate this limit we employ a non-equilibrium Green's function simulator (NEMO) de-

veloped at Texas Instruments Inc.⁶ In Fig. 4, measured and simulated IV characteristics are illustrated for a $Si/SiO_2/Al$ MOS capacitors with different oxide thicknesses. The simulations employ a two-band model with Hartree self-consistent potentials and include injection from both bound and continuum states.⁷ The oxide effective mass ($m_{SiO_2}^* = 0.42m_0$) has been extracted by Brar et. al.⁸ Oxide thicknesses are extracted from CV measurements using a multi-band Hartree self-consistent simulator.⁷

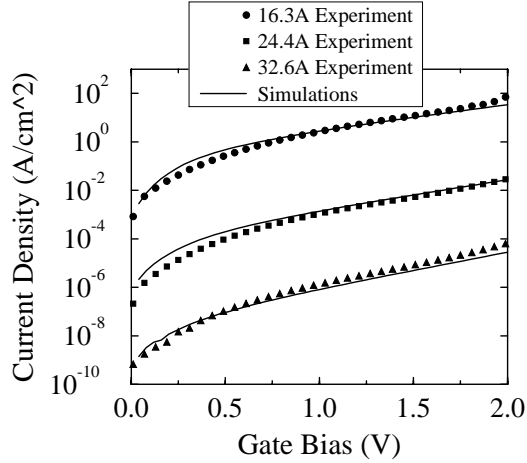


FIG. 4. Current density vs. gate bias for $Si/SiO_2/Al$ MOS capacitors with different oxide thicknesses. Symbols represent data, lines are simulation. The substrate is n-type with a concentration of $1 \times 10^{18} cm^{-3}$.

In Fig. 5, simulated current density vs. oxide thickness at 1.0V is shown. For the purpose of approximating a scaling limit for oxide thickness we shall assume a limit of $1nA$ per micron of gate width as an allowable upper bound for leakage current. For a gate length of $0.1\mu m$, this corresponds to a tunneling current density of $1A/cm^2$ (assuming homogeneous flow through the gate insulator). According to the curve in Fig. 5, this corresponds to an oxide thickness of approximately 17\AA . One can obtain thinner effective oxide thicknesses by employing a gate dielectric with a higher permittivity than SiO_2 . Dielectrics such as TiO_2 ($\epsilon \approx 30$)⁹ and Ta_2O_5 ($\epsilon \approx 25$)¹⁰ have been suggested for gate insulators. However, it is not yet clear if these dielectrics can form high quality interfaces required to obtain high channel mobilities. The use of nitrided oxides has also been studied.^{11,12} These films provide high channel mobilities as well as a marginal reduction in the effective gate oxide thickness.

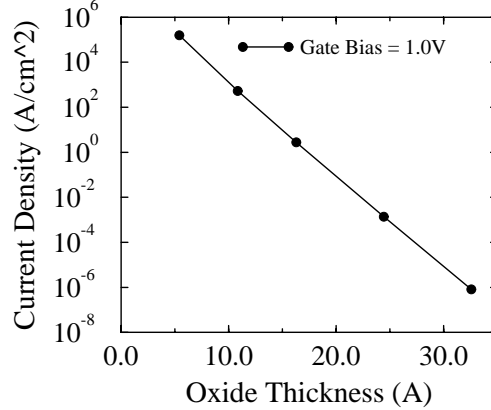


FIG. 5. Simulated direct tunneling current density vs. gate oxide thickness for a $Si/SiO_2/Al$ capacitor with $V_{GB} = 1.0V$.

IV. SIZE QUANTIZATION

Another effect that reduces the performance of MOS devices is energy quantization due to the inversion layer confining potential. This results in a reduction in the density of states in the channel and moves the channel charge centroid away from the gate.¹³ As illustrated in Fig. 6 this reduces the inversion capacitance (and thus the drive current) of the MOS device. For an oxide thickness of 20\AA , quantization results in a $\approx 12\%$ reduction in inversion capacitance. Unfortunately there is not much one can do to reduce or eliminate this effect.

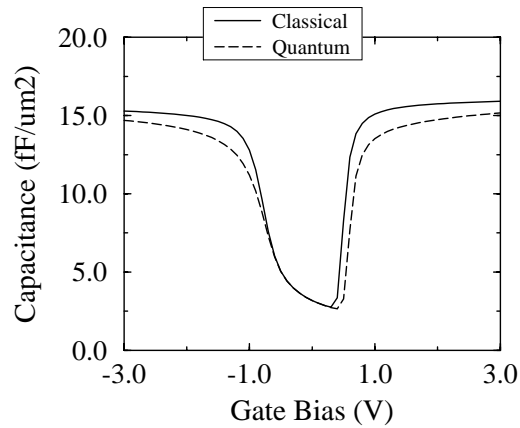


FIG. 6. Comparison between quantum and classical simulation of an MOS capacitance-voltage characteristic. Oxide thickness is 20\AA , channel doping is $N_A = 10^{17}$, and the gate material is TiN. The inversion capacitance predicted quantum mechanically is 12% less than the classical prediction.

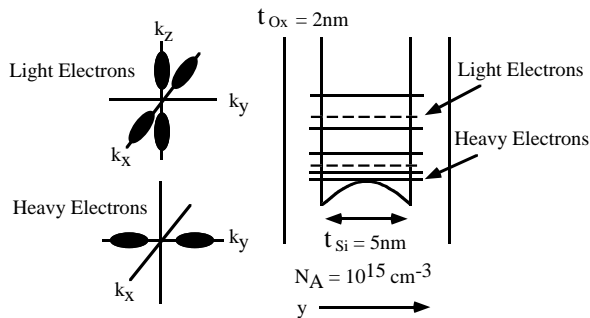


FIG. 7. Conduction band edge and bound state energies for a double gate capacitor. For small silicon film thicknesses, size quantization is enhanced. The density of states and therefore the drive current is subsequently reduced.

We shall now investigate the size quantization effect on the double gate structure. It was shown in section II that this structure provides a solution to the dopant fluctuation problem for thin silicon films. We would like to know at what film thicknesses does size quantization begin to seriously degrade device performance. As shown in Fig. 7, for very thin silicon films the confining potential is determined by the oxide barriers which results in enhanced quantization. One would therefore expect that for thin enough films, degradation due to quantization would also become enhanced.

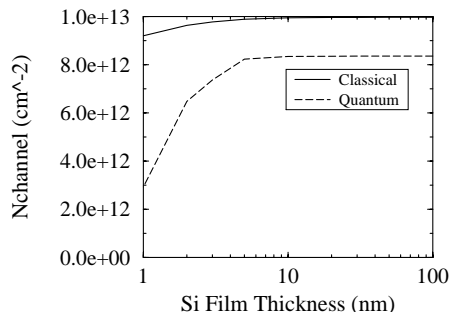


FIG. 8. Double gate capacitor (channel) inversion charge vs. silicon film thickness. Size quantization effects become more pronounced for film thicknesses below ≈ 5 nm. Channel doping is $N_A = 10^{15}$, $t_{ox} = 20$ Å, gate material is TiN, and $V_{gs} = 1.0$ V.

To determine the onset of this enhancement we calculate the inversion charge in a double gate capacitor both classically and quantum mechanically as a function of silicon film thickness. An efficient multi-band Hartree self-consistent Poisson solver is used for the calculations.⁷ In Fig. 8 it is apparent that for Si film thicknesses below ≈ 5 nm quantization begins to seriously reduce the amount of charge

in the channel. As mentioned previously, a ratio between the gate length and the silicon film thickness of approximately four must be maintained in order to avoid punchthrough. This suggests that a double gate structure impervious to dopant fluctuation effects should be scalable to a gate length of about 20nm before quantum effects seriously degrade device performance.

V. CONCLUSIONS

We have investigated the effects of dopant fluctuations, tunneling and size quantization on sub-0.1μm CMOS. It was shown that dopant fluctuation effects can be suppressed with a thin intrinsic layer in bulk devices and eliminated using a double gate device. Direct tunneling current places the scaling limit of SiO_2 gate insulator thickness at approximately 17Å for $L_G = 0.1\mu m$ devices. Alternate gate dielectrics with higher permittivity offer possible methods to push this limit further. Size quantization effects were shown to not severely degrade the performance of the double gate structure until the silicon film thickness is reduced to below 5nm. This suggests that size quantization limits the scaling of this structure to gate lengths of about 20nm.

- ¹ Y. Taur et. al., IBM J. Res. Develop., v.39, p.245 (1995)
- ² V. De, X. Tang, J. Meindl, 54th Dev. Res. Conf. Dig., p. 114 (1996).
- ³ H. Wong and Y. Taur, IEDM Tech. Dig., p.705-708 (1993)
- ⁴ T. Tosaka, K. Suzuki, T. Sugii, IEEE Elec. Dev. Lett., v.15, p.466 (1994)
- ⁵ Unpublished numerical simulations indicate that an aspect ratio of approximately 4:1 between gate length and Si film thickness is required to screen drain from source with no dopant ions present.
- ⁶ G. Klimeck et. al., Appl. Phys. Lett., v.67, p.2539 (1995)
- ⁷ C. Bowen et. al., IEDM Tech. Dig., (1997)
- ⁸ B. Brar, G. Wilks and A. Seabaugh, Appl. Phys. Lett., v.69, p.2728 (1996)
- ⁹ S. Campbell et. al., 54th Dev. Res. Conf. Dig., p.106 (1996)
- ¹⁰ C. Fu et. al., J. Appl. Phys., v.81, p.6911 (1997)
- ¹¹ S. Hattangady et. al., IEDM Tech. Dig., p.495 (1996)
- ¹² X. Wang et. al., 1995 Symposium on VLSI Technology. Digest of Technical Papers, p.109 (1995)
- ¹³ S. Jallepalli et. al., IEEE Trans. Electron Devices, v.44, p.297 (1997)